AD-A247 669
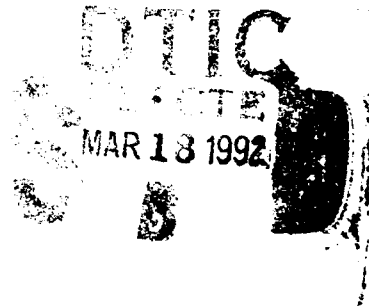
LAPSES IN ALERTNESS: BRAIN-EVOKED RESPONSES

TO TASK-IRRELEVANT AUDITORY PROBES

S. Makeig

F. S. Elliott

M. Inlow

D. A. Kobus

DTIC
MAR 18 1992

Report No. 90-39

92-06883

LAPSES IN ALERTNESS: BRAIN-EVOKED RESPONSES
TO TASK-IRRELEVANT AUDITORY PROBES

Scott Makeig[1]

F. Scot Elliott[1]

Mark Inlow[2]

David A. Kobus[1]

1. Naval Health Research Center, San Diego, CA

2. Geo-Centers, Inc., Fort Washington, MD

Summary

Thirteen subjects participated in an auditory simulation of a passive sonar target detection environment. Targets were 300 ms noise bursts presented at near threshold levels in a noise background at a mean rate of 10 per minute. Task-irrelevant probe tones were also presented at inter-stimulus intervals of 2-4 seconds. Each subject participated in two 28 minute test sessions, pressing a button whenever they detected a noise target. Prominent minute-scale fluctuations in performance (computed as changes in local error rate using a 32-s moving window) occurred in many of the sessions. Evoked responses to the irrelevant probe tones in thirteen runs with highest number of performance lapses were sorted by current local error rate and smoothed using a moving-average. The amplitude of the grand mean N2 response to the irrelevant probe tones increased monotonically with error rate. Averaged evoked responses to relatively frequent, task-irrelevant probe tones appear to allow an accurate estimate of level of alertness if adequate numbers of trials are available.

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	

By

Distribution/

Availability Codes		
Dist	Avail and/or Special	
A-1		

2

Introduction

Since the experiments of Mackworth (1948), decrements in performance on continuous supra-threshold detection tasks have been referred to as "vigilance decrements." Mackworth's results supported his recommendation that operators performing tasks which require sustained vigilance should be relieved after 30 minutes on task. However, in such experiments mean trends in performance across sessions may bear little resemblance to fluctuations in responsivity in sessions. After monotonous test sessions, subjects may recall having become drowsy but still report having experienced only brief periods of daydreaming or absentmindedness and deny having dozed at all, even after having ceased responding altogether for several minutes. Knowledge of mean trends in performance, however, does not allow accurate prediction of when an individual operator or pilot is unable to detect and respond to inconspicuous but important signals or events. Nonetheless, relatively little research has been concerned with quantifying and studying fluctuations in alertness on a minute-to-minute basis. This is in part because, out of desire to simulate real-life work environments, most vigilance research has used target presentation rates too low to define such fluctuations behaviorally.

Since Loomis' original observations of electroencephalographic (EEG) changes in sleep (Loomis, Harvey, and Hobart, 1937), researchers have imagined the possibility of developing an electrophysiological monitor of alertness. But while relatively successful methods have been developed for automatically categorizing sleep stages using combined EEG and electrooculogram (EOG) measures (see Goeller and Sinton, 1989), the problem of tracking the electrophysiological signs of loss of alertness due to transition to sleep has proven difficult.

In such research, most attention has been given to changes in the appearance (Santamaria and Chiappa, 1987), topography (Ulrich and Frick, 1986), or spectrum (Townsend and Johnson, 1979) of the spontaneous EEG during transition to sleep as defined by EEG criteria. However, several reports have studied electrophysiological correlates of decline in alertness using continuous performance measures. Beatty, Greenberg, Deibler, and O'Hanlon (1974) reported that periods of lowered performance on a visual task were predicted by lowered electrocortical activation as indexed by a simple measure of occipital theta band EEG amplitude.

Matousek and Petersen (1983), using a linear combination of over thirty EEG band and band-ratio amplitude measures, were able to reproduce the visual classification of EEG epochs into awake and Stage I sleep states by visual inspection. Belyavin and Wright (1987) reported that changes in amplitudes in theta (4-7 Hz) and beta (14-21 Hz) EEG bands could be used to predict changes in performance on a simple vigilance task, while delta (1-3 Hz) and beta band levels were best predictors of fluctuations in performance on a more difficult discrimination task. Torsvall and Akerstedt (1988) observed that slow eye movements and a large increase in alpha EEG precede dozing off during performance of a simple behavioral task. Ogilvie, Simons, Kuderian, MacDonald, and Rustenburg (1991), using an auditory response task at the beginning of all-night sleep sessions, have recently claimed that power in all EEG bands increases when sleep-related lapses first occur.

Most other reports have studied the transition from waking EEG to sleeping EEG using EEG rather than behavioral criteria (see, for example, Fruhstorfer and Bergstrom, 1969; Penzel and Petzold, 1989). However, all authors agree that across subjects, recognized EEG correlates of drowsiness -- including alpha amplitude increase and frontal spread, minute-scale spectral variability, slow horizontal eye movements, appearance of slow wave activity and sleep spindles -- may not be strongly correlated with changes in performance. Further, across subjects the appearance of these signs varies substantially (Santamaria and Chiappa, 1987), limiting the potential success of subject-independent linear prediction algorithms.

Another approach to alertness monitoring involves measures of event-related potentials (ERPs). Several endogenous ERP features linked to cognitive processes, foremost among them the P300, are well known to index the allocation of attention to the evoking stimulus, and therefore fade or disappear during at or near sleep onset (reviewed in Kramer, 1991). However, cognitive evoked response methods in general, and P300 recordings in particular, require that the moments of occurrence of task-relevant events must be known precisely. Unfortunately, in many settings the moments of occurrence of events relevant to the performance of the operator's task can neither be known in advance nor detected automatically. In these environments, therefore, measures of the P300 evoked following target-relevant signals, or other endogenous potentials evoked by task-relevant signals cannot be used to monitor

4

alertness unless artificial secondary tasks and task-related stimuli are introduced. But these may tend to focus the operator's attention away from their primary task, thus reducing rather than enhancing operator performance.

In such situations, it would obviously be preferable to use brain responses evoked by task-irrelevant signals, if they were known to signal the early onset of performance decrements. In this approach, stimuli irrelevant to the operator's task would be delivered to operators periodically to probe the ability of the central nervous system to respond to sensory stimulation. Averaged event-related potentials (ERPs) to task-irrelevant probes have also been shown to change profoundly in sleep. Early auditory ERP studies showed that a large N2 wave (peaking circa 350 ms) emerges at sleep onset and dominates sleeping responses (Weitzman and Kremen, 1965; Williams, Tepas, and Morlock, 1962). Ornitz, Ritvo, Carr, La Franchi and Walter (1967) reported that N2 response amplitude is greatest within 5-10 minutes of sleep onset (defined by the first appearance of sleep spindles in the EEG).

Fruhstorfer and Bergstrom (1969) attempted to measure changes in the auditory evoked response during sleep transitions in more detail. They used a midline bipolar electrode montage and presented clicks at 8-20 s intervals to subjects instructed to fall asleep and pay no attention to the sounds they heard. By visually inspecting the EEG traces, they then categorized each response epoch into one of nine stages of electroencephalographic vigilance according to the criteria of Roth (1961) and Bente (1964). They reported that the click-evoked responses peaks N1 and P2 decline in amplitude with decrease in EEG vigilance (i.e., appearance of EEG changes associated with transition to sleep), and that a new N2 potential, with a shorter latency and more posterior scalp distribution, appears in the later sub-stages. Their figures show that for two subjects, this sleep N2 first appeared in their stage B2, characterized by first appearance of 5-7 Hz slow waves. For the other four subjects, it appeared in stage C, characterized by mixed 3-7 Hz and 12-14 Hz rhythms. These stages roughly correspond to stage I sleep according to standard terminology (Rechtschaffen and Kales, 1968). However, Fruhstorfer and Bergstrom (1969) did not report performance rates in their EEG sleep stages. Accordingly, it remains of interest whether changes in the N2 and other features of the task-irrelevant evoked response occur with early, intermittent decline in performance, or whether

they occur only after subjects cease responding completely.

Some evidence on this point has recently been presented by Ogilvie et al. (1991), who presented relatively long duration (<= 5s), low level (27-30 dBSL) target tones at relatively long intervals (mean ISI 17.5s) during the first parts of all-night sleep sessions in which subjects were instructed to respond to detected tones while at the same time allowing themselves to fall asleep. Examination of evoked responses to first to fourth successive tones not responded to revealed that the typical auditory sleep response features (including the sleep N2 and later event-related negativities they identify as K-complexes) were also visible in part in averages of responses to those tones to which their subjects responded most slowly. However, Ogilvie et al. did not attempt to quantify the actual time-structure of fluctuations in alertness in their subjects, only noting that transitions could be shorter than the sampling rate (40 s per estimate) used in normal sleep staging.

In the present experiment, we use relatively frequent (20/min) supra-threshold task-irrelevant auditory probe tones to evoke brain potentials whose features correlate with rises and falls in error rate on a supra-threshold detection task requiring sustained attention. The paradigm we use falls neither into the mold of most classical vigilance experiments, in which responses to infrequent target events are studied, nor does it follow the signal detection paradigm, which normally employs near-threshold signals and assumes that operator state is not continuously fluctuating. Our experiment is also unlike most sleep studies in which subjects are instructed to fall asleep. In our case, although the experimental setting was conducive to drowsiness, the subjects were instructed to attempt to perform the task continuously through half-hour simulated sonar watches. Elsewhere (Makeig and Inlow, unpublished), we study in detail the spectral structure of these fluctuations in alertness and explore their relationship to simultaneous changes in the EEG spectrum. Our aim here is to quantify the temporal characteristics of fluctuations in performance on the task, and then to show that several features of sensory evoked responses to irrelevant auditory probes correlate with concurrent performance level as drowsiness or inattention overtakes the operator and behavioral responses become sporadic. In particular we find that averaged responses to probes occurring just before (1) targets responded to (Hits) and (2) targets not responded to (Lapses) are linearly separable.

Finally, we discuss possible applications of this research to methods of objectively monitoring the current vigilance level of operators.

Method

Subjects. Thirteen males participated in an auditory simulation of a passive sonar target detection task. Of these, nine were prospective students in a Navy sonar course, three were sonar instructors, and one was from the laboratory staff. Ages ranged from 18 to 34 (mean 24 years). All had passed standard Navy hearing tests.

Stimuli. Sound synthesis and data collection were controlled by a Concurrent Realtime Unix computer system using a 12-bit D/A converter sampling at 50 kHz. The three stimulus streams used in the experiment are shown schematically in Figure 1. All stimuli were presented binaurally through headphones mounted in isolating cuffs in a white noise background at 62 dB nHL. Task-irrelevant auditory probe tones of two frequencies (568 and 1098 Hz) were presented in random order with inter-stimulus intervals between 2-4 s at 72 dB nHL. Probe tones were 50 ms in duration, with rise and fall times of 10 ms. To explore use of response features associated with stimulus novelty, the high tone was presented more frequently (80%) than the low tone (20%). The two stimuli will therefore be referred to below as Frequent and Rare respectively.

Target noise bursts were 300 ms in duration with long rise and fall times of 150 and 110 ms respectively. Targets occurred in 50% of the 2-4 s inter-probe intervals, giving a mean target presentation rate of 10 per minute. Target intensity was set at 6 dB above its relative threshold in the noise. This intensity was high enough to produce initial performance levels at or near ceiling, but was low enough not to startle subjects or delay onset of alertness decrements.

Steady-state click probe stimuli were also presented continuously through the experimental sessions at a rate of 39 Hz and an intensity of 63 dB nHL. Within the noise background, the click train was perceptible but not intrusive. It was presented to evoke a stable steady-state
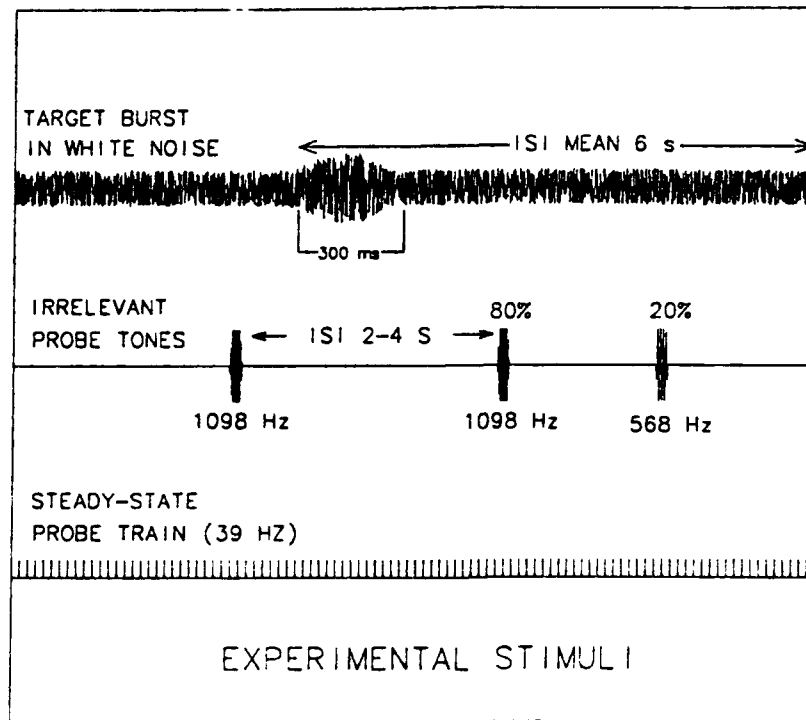
```
┌──────────────────────────────────────────────────────────────┐
│                                                                │
│  TARGET BURST                                                  │
│  IN WHITE NOISE      ◄──────────── ISI MEAN 6 s ────────────►  │
│                                                                │
│  [waveform]                                                    │
│                      └─300 ms─┘                                │
│                                                                │
│  IRRELEVANT                          80%         20%           │
│  PROBE TONES    █ ◄─ ISI  2-4 S ─►█            █               │
│                                                                │
│               1098 Hz            1098 Hz     568 Hz            │
│                                                                │
│  STEADY-STATE                                                  │
│  PROBE TRAIN (39 HZ)                                           │
│  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  │
│                                                                │
│           EXPERIMENTAL  STIMULI                                │
│                                                                │
└──────────────────────────────────────────────────────────────┘
```

Figure 1. Schematic view of the three stimulus streams in the Experiment. Note that durations and amplitudes are not to scale.

response (SSR) which will be discussed in another report also dealing with changes in the EEG spectrum.

Procedure. Each subject participated in two simulated work sessions of 28 minutes. Subjects sat in a comfortable chair with eyes closed, their right index finger resting on a response button. They were instructed to press the button as soon as possible each time they detected a noise burst, and to ignore the probe tones.

Physiological Recording. Data from all sessions were continuously recorded to disk for off-line analysis. EEG and EOG signals were amplified 50K times with a 0.1-100 Hz bandwidth through Grass EEG amplifiers, then multiplexed with button press information and converted

to 12-bit digital format at a sampling rate of 312.5 Hz per channel. EEG was collected from 13 scalp locations of the International 10-20 system. An ECI Electro-Cap provided standard-ized placement of Ag/AgCl electrodes at 13 sites (Fpz, F3, Fz, F4, ⊤3, C3, Cz, C4, T4, P3, Pz, P4, Oz) referred to the right mastoid. A left-to-right mastoid electrode channel was also collected and data was re-referenced to digitally-linked mastoids during averaging. Periocular electrodes were used to record electrical potentials generated by voluntary and involuntary movements of the eyes during EEG recording. These consisted of one pair of electrodes placed horizontally at the outer canthii, and a second pair placed one inch above and below the left and right eyes respectively. Electrical impedances at all electrode sites were less than five kOhms.

Analysis. Responses to the target noise bursts were divided into three categories. If the subject pressed the button within a time window of 150-2000 ms following the target onset, the stimulus was designated a Hit. If no such response was made, it was labeled a Lapse. Stimuli followed by inappropriate responses were labeled as Errors or False Alarms, but these were so few in number (<6%) that stable averages of responses to them could not be formed, and they will not be discussed further. Averaged evoked responses were computed for each stimulus category. To prevent eye blinks or muscle potentials from contaminating recordings, EEG epochs were excluded during averaging if potentials at any site exceeded + 90 uV. To remove the SSR produced by the time-locked 39/Hz click stimuli, before analysis the averaged data were lowpass filtered using a finite-impulse response (FIR) filter with a cutoff frequency near 32 Hz, and a slope of 12 dB per octave. To construct a local measure of performance, the times of occurrence and response status of target stimuli were averaged using a moving-window averaging algorithm which maintained a fixed width (in seconds) and step size (1.6 s).
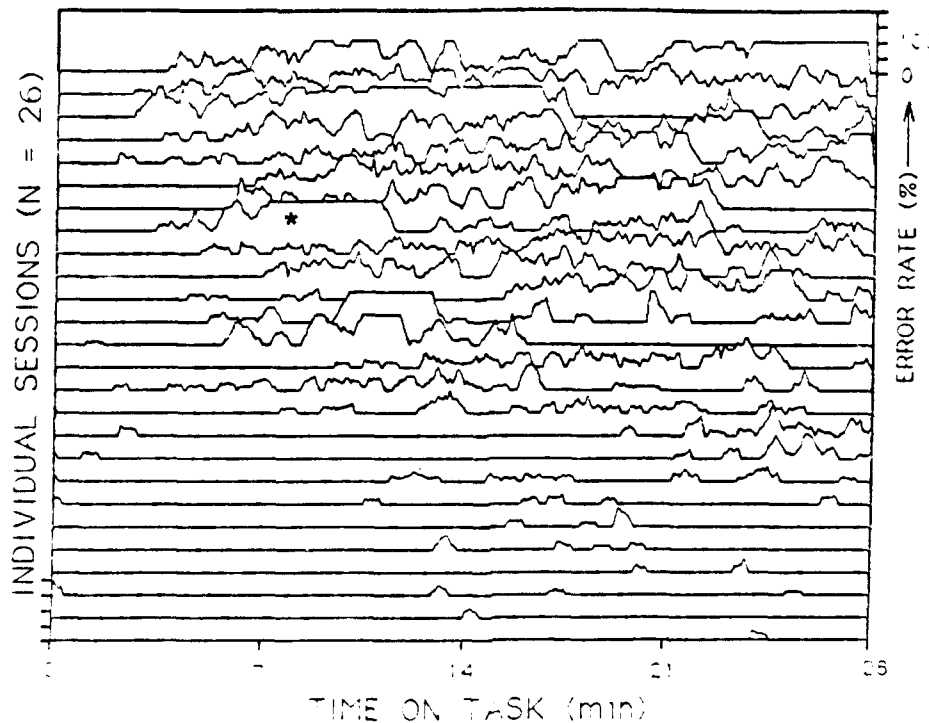
Results

Figure 2. Fluctuations in local error rate for all 26 sessions computed using a 32s moving window).

Performance

Local error rate. The variability of individual session records is illustrated in Figure 2, which shows fluctuations in local error rate (computed using a 32 s moving window) for each of the 26 sessions. As the Figure shows, in these sessions performance varies from relatively poor to near perfect. In sor.e. intermittent lapses begin early in the session and continue throughout, while in one session (marked with an asterisk), only four minutes of good performance occur before a transition to a period of nearly five minutes during which the subject

did not respond at all. After this prolonged absence, performance returns to near perfect, marked only by intermittent lapses every few minutes. At the end of this session, the subject reported no memory of failing to respond.
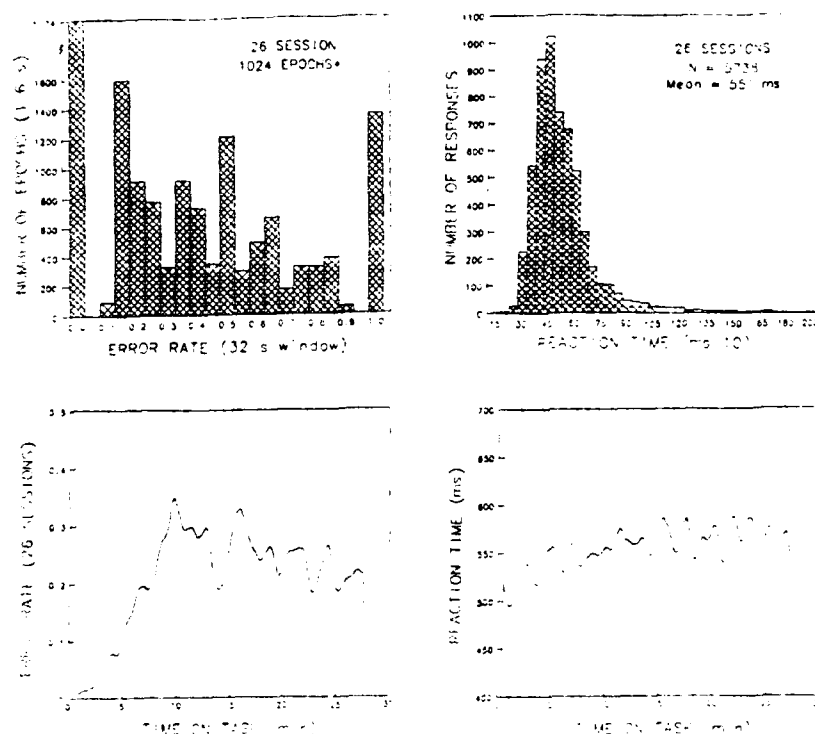

Figure 3. Mean performance measures for data from all 26 sessions. (a) Error rate histogram. (b) Reaction time histogram. (c) Error rate as a function of time on task. (d) Reaction time as a function of time on task. Both (c) and (d) were smoothed using a circa 50 s window.

Figures 3a and 3b show the performance and reaction time (RT) histograms across the 26 sessions, each divided into 1024 epochs of 1.6 seconds each. Local error rate (in a 32 s moving window) was non-zero in 35% of the epochs; in 5% error rate was 100%. Figure 3c

shows performance as a function of time on task averaged over all 26 sessions using a 105 s moving average. This figure resembles classic results of vigilance decrement (Mackworth, 1970): performance remains optimum for about three minutes, then error rate rises, in this experiment reaching a plateau at 10 minutes into the task. The mean error rate maximum at 10 minutes is caused by a few subjects who at that point ceased responding for some minutes, then resumed responding. However, it is clear from Figure 2 that few of the individual session records closely resemble the behavior of the mean shown in Figure 3c. Figure 3d shows mean RT as a function of time on task (also averaged using a 105 s moving window). Mean RT also rises during the first part of the sessions, then stabilizes.
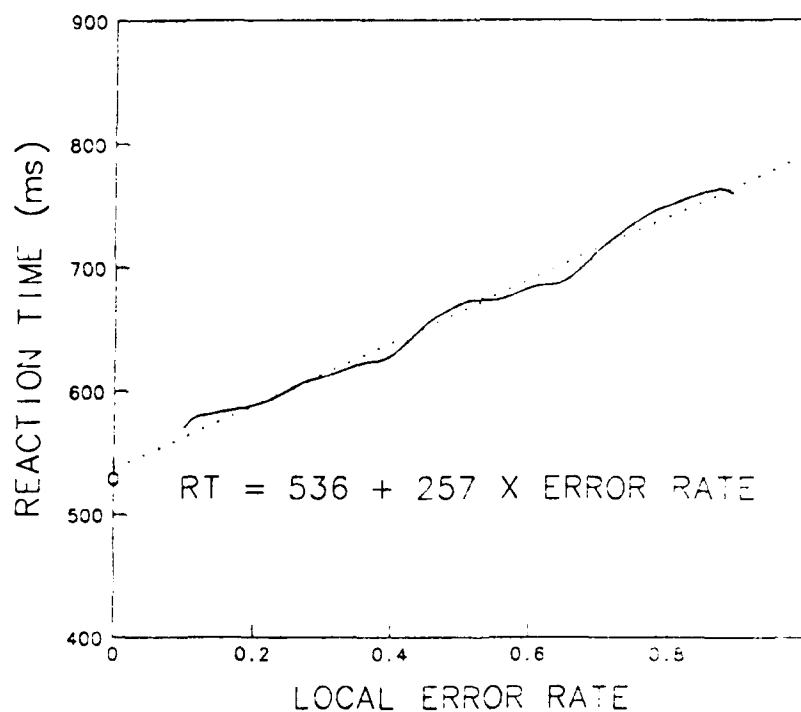

$$RT = 536 + 257 \times ERROR\ RATE$$

Figure 4. Mean reaction time averaged as a function of local error rate using a 10% moving window.

Reaction time versus error rate. Figure 4 plots mean reaction time by local error rate. The plot is remarkably linear across almost the entire range of error rates, the regressed value increasing by 257 ms from 0% to 100% errors. This figure suggests that changes in the local error rate measure are as meaningful a measure of performance as changes in mean RT.

Evoked Responses

Probe Responses. As expected, averaged responses to the noise target Hits contained a prominent P300 maximum at Pz, while those to target Lapses did not. The target response data will be presented elsewhere. Figure 5 shows grand mean ERPs to the Frequent and Rare task-irrelevant probes at the vertex, where these potentials are largest, selectively averaged over stimuli which immediately preceded correctly detected target bursts (Hits) or undetected targets (Lapses). Note that neither the pre-Lapse nor pre-Hit responses to the Rare irrelevant probes contain a P3 peak, confirming that the subjects allocated little or no attention to the probe.

Task-irrelevant probe responses to frequent probe stimuli preceding target Hits and Lapses differ in at least three ways (see Figure 5). Before Lapses, the prominent P2 and N2 peaks are larger, and the N1 deflection is smaller than prior to Hits. The uniformity of these changes across subjects was studied by means of analysis of variance on the 14 sessions from 10 subjects including at least 40 Lapses each.

Figure 6 shows the difference (pre-Lapse minus pre-Hit) between Frequent N2 probe response peak amplitudes as a function of the number of sums in the pre-Lapse evoked response. In sessions with fewer lapses, the N2 difference was more variable. This may have been because the smaller number of sums made the signal-to-noise ratio of the pre-Lapse ERP small, or else because occasional Lapses occurring during the relatively error-free sessions may not have been associated with sleep transitions and associated appearance of the N2, but rather were due to inattention or rare random failures of signal detection.

Scalp distribution. As Frequent probe response averages contained many more sums than Rare probe responses, and were therefore more reliable, an analysis of variance was performed on responses to the Frequent probe stimuli alone. At frontal and central channels the

POTENTIAL (uV)

10
8
6
4
2
0
-2
-4
-6
-8
-10

Rare→Lapse

Freq→Lapse

IRRELEVANT
PROBE EPS
N = 26 SESSIONS

Freq→Hit

Rare→Hit

→Hit

Freq→Lapse
Freq→Hit
Rare→Lapse
Rare→Hit

→Lapse

Cz - DLM

-400 -200   0   200 400 600 800 1000 1200 1400
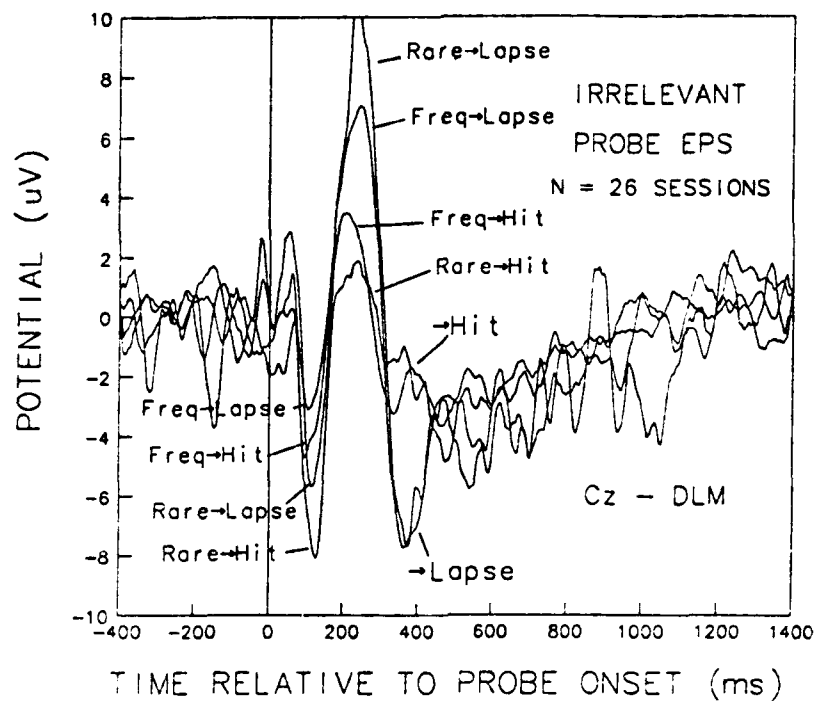
TIME RELATIVE TO PROBE ONSET (ms)

Figure 5. Irrelevant probe evoked responses. Responses to Rare and Frequent probe stimuli occurring before detected (Hit) and undetected (Lapse) targets.

P1-N1 difference was smaller pre-Lapse than pre-Hit ($F(1,9)$ at all channels > 11.2; p<.01). At left temporal sites, N2 amplitude was significantly larger preceding Lapses than preceding Hits ($F(1,9)>9.25$; p<.015), and P2-N2 peak difference was also larger at the three sites Cz, Fz, and F3 ($F(1,9)>11.8$; p<.01). Topographic maps of significance levels for these effects are shown in Figure 7. This figure was constructed by performing ANOVAs separately on data from each channel, converting the F-values to probability levels, and mapping the results using weighted nearest-neighbor interpolation.
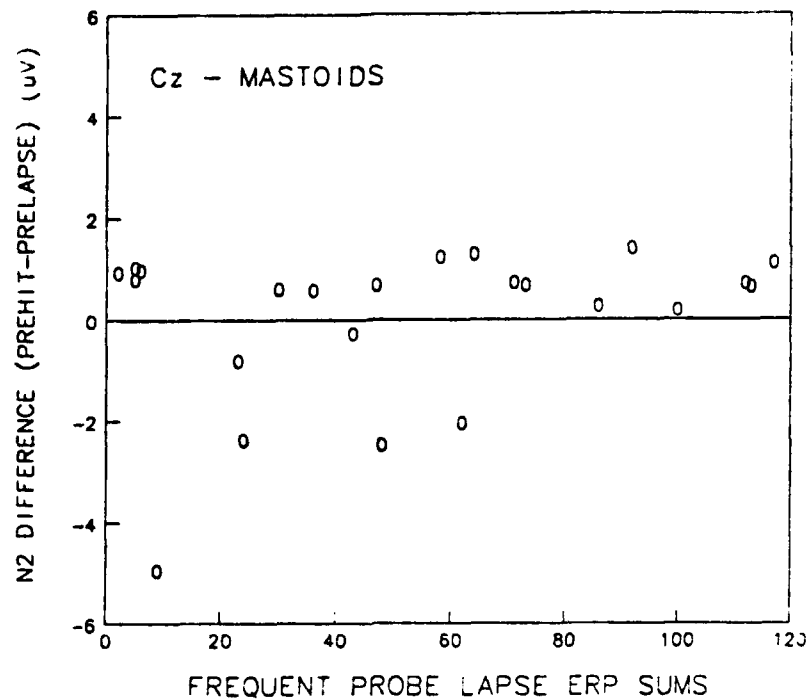
Figure 6. Difference in N2 amplitude versus performance. Abscissa is the number of sums included in the Frequent probe pre-Lapse evoked response. Ordinate is the difference in N2-to-baseline peak amplitude between pre-Hit and pre-Lapse responses.

Discriminant Analysis

As a first test of the feasibility of using irrelevant probe ERPs to estimate behavioral alertness, we attempted to classify Rare and Frequent averaged probe responses for sites Fz, Cz, C3, and C4 from the 15 sessions with highest number of Lapses, categorizing them into two groups: those preceding detected targets (pre-Hits) and those preceding targets missed (pre-Lapses). Values for peak amplitude and latency of peaks P1, N1, P2, and N2, and peak amplitude differences P1-N1 and P2-N2 at all electrode sites were used for initial classification. The classification functions were computed using the stepwise discriminant

P1-N1 DIFFERENCE   N2 COMPONENT

BLACK p<.01
WHITE p>.05

P2-N2 DIFFERENCE
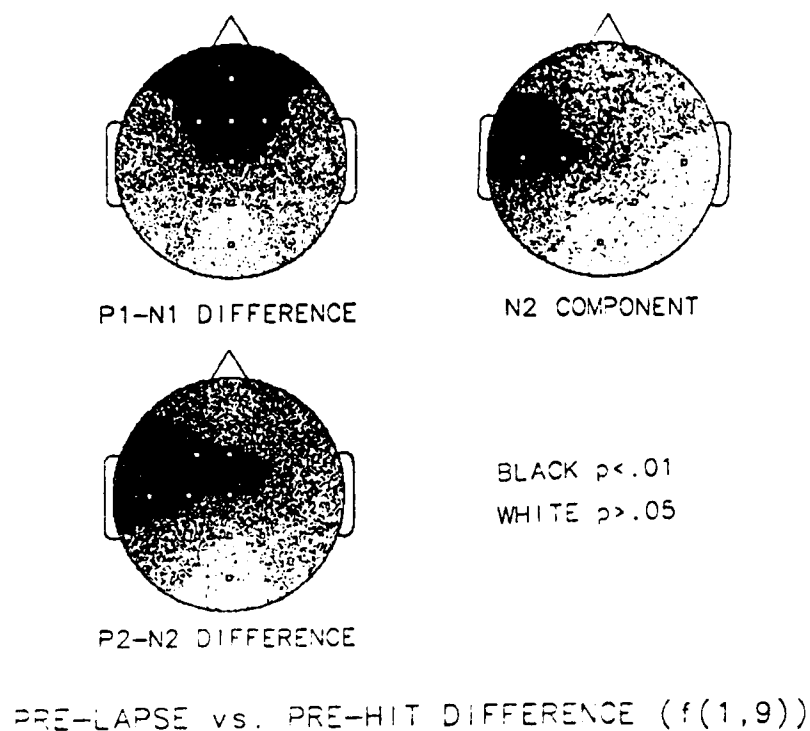
PRE-LAPSE vs. PRE-HIT DIFFERENCE (f(1,9))

Figure 7. Significance maps for difference in Frequent probe P1-N1, N2 and P2-N2 amplitudes between pre-Lapse and pre-Hit responses. Scale: black, $p<.01$; grey, p $<.05$; white, p>.05

analysis program BMDP 7M. Though the 15 sessions used were from 10 subjects, we computed statistics on success of discrimination over sessions, without regard to subject in order to explore which ERP sites and features could most reliably discriminate our two classes of data from one another. Initially we used all of the data [13 channels * 12 measures] as the training set. Since these variables were highly correlated with one another, an exploratory mode of analysis was used in order to rank order measures in order of best discrimination. Responses were classified as pre-Hit or pre-Lapse by assigning them to the nearest group mean using Mahalanobis distance. The five measures chosen first by the initial discriminant

analysis are given in Table 1 in the order selected.

TABLE 1. ERP MEASURES SELECTED FOR PRE-HIT
VERSUS PRE-LAPSE DISCRIMINATION

Step	Measure	F-to-Enter
1.	Frequent N2 latency at Fz	11.9
2.	Frequent P1-N1 amplitude difference at Cz	11.5
3.	Frequent N2 amplitude at C3	17.9
4.	Rare N2 amplitude at Cz	8.2
5.	Frequent N1 latency at C4	7.7
6.	Frequent P1-N1 amplitude difference at Fz	5.3
7.	Rare P2-N2 amplitude difference at C4	4.3

Using these seven variables, a 100% discrimination between pre-Hit and pre-Lapse responses was achieved. As a first test of the reliability of discrimination, a jackknife procedure was then used in which discrimination criteria based on each combination of 14 sessions were used to classify the responses from the 15th session. Only one classification error occurred, though variables chosen varied slightly.

TABLE 2. CLASSIFICATION MATRICES FOR 5 VARIABLE DISCRIMINATION

Test	Response	Percent Correct	Classification	
			pre-Lapse	pre-Hit
training	pre-Lapse	100.0	15	0
	pre-Hit	100.0	0	15
jackknife	pre-Lapse	86.7	13	2
	pre-Hit	100.0	0	15

Since our sample size is small and our measures highly correlated, we tried to make our variable selection more robust by running the same stepwise discriminant analyses using eight random subsets of 12 of the 15 sessions. As before, the Frequent probe responses dominated, with only one Rare probe response measure being included. The measures selected were the original top five except for "Frequent N2 latency at C4," which replaced the same variable at site Fz. This substitution demonstrates the difficulty of choosing a "best" set of classification variables when the variables are highly correlated. Table 2 shows the classification matrices derived for the five selected variables. The top half presents the results from using the initial training set as the test (validation) set, and the bottom half gives the results of a standard "jackknife" procedure in which each observation is classified using classification functions computed from the remaining observations. Clearly the pre-Hit and pre-Lapse ERPs differ enough for successful classification of operator performance.

TABLE 3. CLASSIFICATION MATRICES FOR 3 VARIABLE DISCRIMINATION

Test	Response	Percent Correct	Classification pre-Lapse	pre-Hit
training	pre-Lapse	80.0	12	3
	pre-Hit	93.3	1	14
jackknife	pre-Lapse	80.0	12	3
	pre-Hit	93.3	1	14

Next, we ran this analysis restricting the variables to the first three selected: "Frequent N2 latency at C4," "Frequent P1-N1 amplitude difference at Cz," and "Frequent N2 amplitude at C3." The resulting jackknife classification matrices (Table 3) were as accurate as the training data results, suggesting that classification functions based on these three measures should generalize well to other sessions. Subsequent to this analysis, we discovered that mechanical problems had interfered with the recording of one of the 15 sessions used in the discriminant analysis. Had data from this session not been used, the classification results would have been

even higher.

Finally, to study how classification performance is degraded by using measures of responses from a single site, we used the same three component measures but restricted them to site Cz. The results (Table 3) were comparable in power to the multi-channel classification, suggesting that in operational use, multi-site recording of auditory responses might not offer significant advantages over single-site recording.

BLACK p<.01

WHITE p>.05

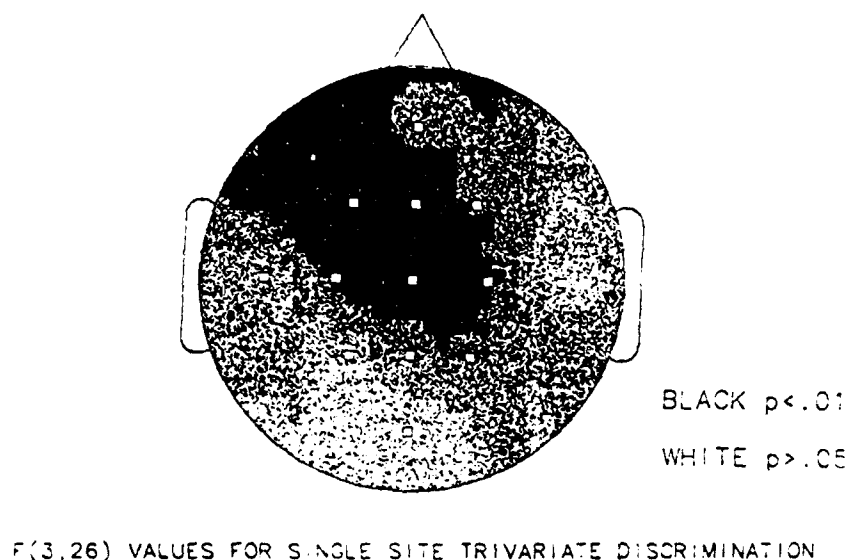F(3,26) VALUES FOR SINGLE SITE TRIVARIATE DISCRIMINATION

Figure 8. Significance map for discriminability of pre-Hit and pre-Lapse ERPs. $F(3,26)$ transform of Wilks' lambda statistic for three peak amplitude and latency measures. (Scale: Black, $p<.01$; grey, $p<.05$; white, $p>.05$).

To determine which site on the scalp showed the most reliable differences between pre-Hit and pre-Lapse responses, we estimated the spatial distribution of the classification

information provided by our top three measures by computing Wilks' lambda for each site, suitably transformed to have an approximate F distribution. In Figure 8 we have mapped the approximate F-values at each recording site. Since the upper 0.001 of the F(3,26) distribution is the lower bound of the Figure scale, it is clear that there is a significant difference between the pre-Lapse and pre-Hit session averages at all sites. The maximum F-value occurs at or near the three sites Cz, Fz, and F3, indicating that for the selected sessions, these measures, in a multivariate sense, attain their most discriminable pre-Lapse to pre-Hit difference in this region. Note that this region is the intersection of the highly significant regions for P1-N1 and P2-N2 differences in Figure 7.
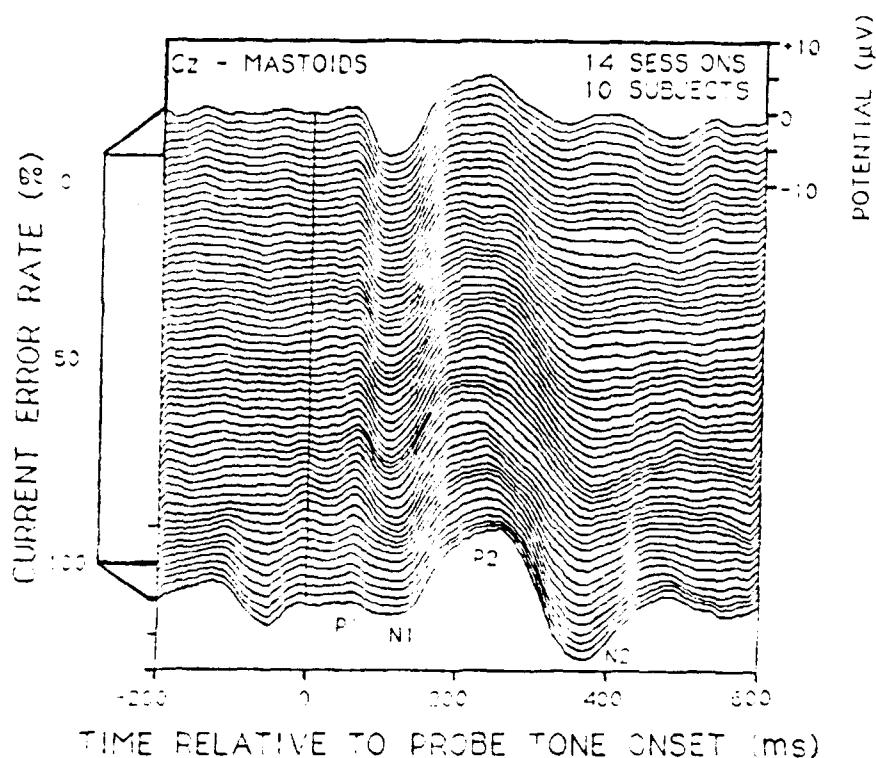

Figure 9. Grand-mean Frequent irrelevant-probe evoked response as a function of local error rate.

Error-sorted Moving Average

To explore the dynamics of the significant changes between pre-hit and pre-Lapse responses, the entire set of single trial responses at Cz were ordered by local error rate computed using a moving window of 32 seconds. A moving average of this reordered data was then constructed using a window width of 30% (error rate) and a step size of 2%. These values were chosen to yield smoothly varying mean estimates. The results of this procedure are shown in Figure 9. The top trace in this figure shows responses centered in error-free epochs; the bottom trace shows responses in no-response epochs. In the figure, as error rate increases, a prominent N2 appears and P1-N1 decreases.
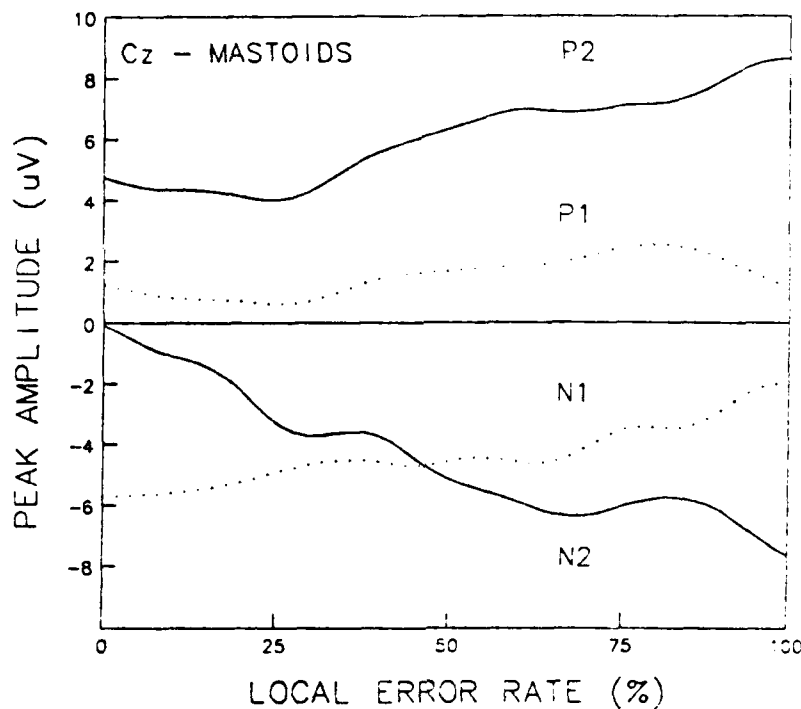

Figure 10. Grand mean Frequent irrelevant probe ERP peak amplitudes versus local error rate. Data from Figure 9.

Figure 10 plots the amplitudes of grand mean P1, N1, P2, and N2 peaks versus error rate. This figure was constructed by determining the latency of the maximum for each peak, and extracting potentials at these latencies from the data of Figure 9. Grand mean N2 amplitude increases from baseline, at 0% errors, to 8 uV at 100% errors, though it is not clear whether there is any increasing trend for midrange error rates from 30% to 70%. Grand mean P2 appears constant until error rate reaches approximately 30%, and then to increase in size as error rate rises further. Meanwhile P1 amplitude remains stable across error rates up to 90%, while the size of the N1 potential decreases nearly monotonically as error rate rises.
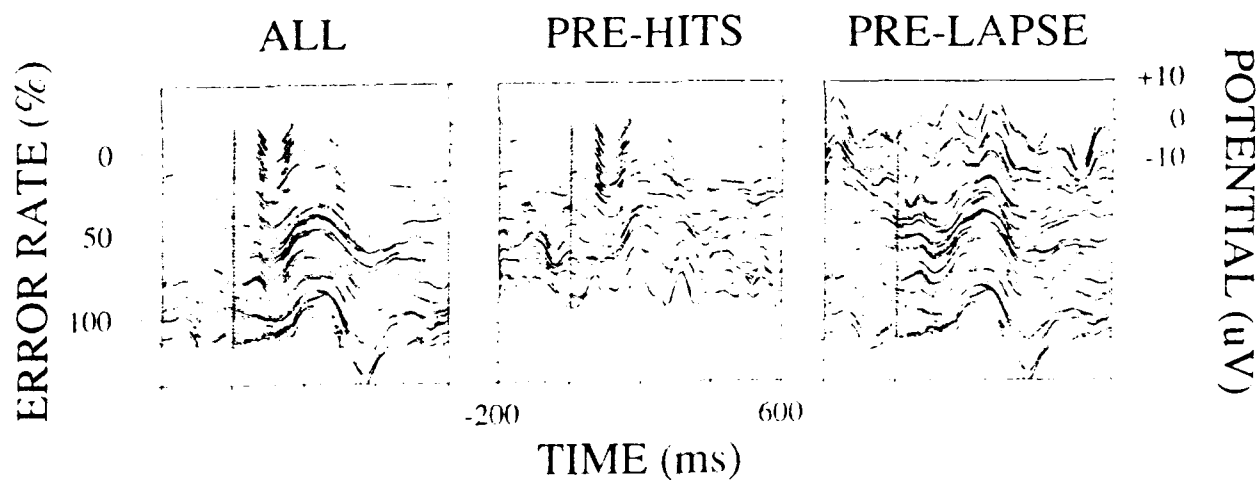

Figure 11. Frequent irrelevant probe ERP as a function of local error rate. Left panel, means of all probes preceding targets. Center panel, means of all probes preceding Hits. Right panel, means of all probes preceding Lapses.

Figure 11 computes the grand mean moving averages of just those Frequent probes which immediately preceded noise targets (with the further condition that at least 600 ms separated the noiseburst and probe tone onsets). In the center panel, the moving average of those probes preceding Hits is shown, and in the right panel, the moving average of probes preceding Lapses. The left panel shows the moving average of both categories. Naturally, this average is dominated by pre-Hit epochs at low error rates, and by pre-Lapse epochs at high error rates, and this is reflected in the form of the evoked response.

Discussion

The present study has shown at least two features of auditory evoked responses to completely task-irrelevant probe tones can index subject alertness without requiring that the subject attend to the probe tones or perform a secondary task. In particular, our results show that the auditory N2 wave (in this study peaking near 380 ms), which is known to become prominent during deeper stages of sleep, first appears just as alertness begins to become sporadic and, across subjects, increases in amplitude monotonically and near-linearly as frequency of performance lapses increases. Emergence of this sleep-N2, signaled by an increased latency for the N2 peak measure, was the measure which best separated pre-Hit from pre-Lapse responses.

Also, as lapses in attention became more frequent, peak N1 amplitude steadily decreased, and this quantity (best estimated in our data as P1-N1 amplitude difference), also reliably differentiated mean pre-Hit from pre-Lapse responses. This result complements the many studies which have shown that attending a stream of sounds enhances N1 amplitude (Hillyard, Hink, Schwent, and Picton, 1973) to sounds in the attended stream.

A primary goal of this research was to explore the possibility that task-irrelevant evoked responses could be used to monitor operator alertness. In ERP research on performance estimation, most attention has been paid to use of the P300 response which is evoked when a target or unanticipated task-relevant event is detected. The P300 may be useful to monitor

attention to, or workload during machine-paced primary or secondary tasks (Kramer, 1991). However, in many actual work environments, recording ERPs to target information may not be possible, and task irrelevant probe responses appear to be a possible replacement.

Our results show that at least three measures of averaged irrelevant-probe ERPs appear to differ substantially enough to enable pre-Hit and pre-Lapse states to be discriminated. However, note that we tested only those sessions in which the pre-Lapse ERPs were averages of at least 40 trials. The averaged responses discriminated in these procedures were then averages of data collected over entire half hour sessions. A method which detected loss of alertness only after many minutes of seriously impaired operator capacity would only partially achieve timely cognitive status monitoring. To be of practical use, a discrimination procedure would have to deal with averages of fewer sums to estimate changes in performance within sessions.

However, our results do suggest that single channel classification performance appears competitive with multi-channel results, a result with practical advantages since single-channel recording could be much easier than multi-channel in the workplace. The optimum site for performance estimation may be in the vertex region (Cz to Fz). While a slight left-sided bias in predictive power appears in our N2 data, response state-by-side interactions were not statistically significant. Also, in this experiment, ERPs to the frequent irrelevant probes provide most of the useful information for classification. Though this might be due to the larger number of frequent probe ERPs available for averaging, it suggests that the use of rare or "oddball" irrelevant probes may not give more robust information about performance than responses to single-frequency irrelevant probes alone.

In our view, an optimal alertness monitoring system would not use probe ERP data alone. The EEG averaged to create evoked responses also contains much information about operator state which may not be reflected directly in the ERP (Townsend and Johnson, 1979; Makeig and Inlow, unpublished). Other physiological phenomena, including heart-rate fluctuations, slow eye movements, and steady-state response amplitude, may all provide convergent evidence of fluctuations in alertness that need not and should not be ignored by designers of such a system.

The statistical methods used to predict performance are also capable of much improvement. The linear discriminant procedure we have used in this report creates classification functions which are optimal only if the variables are normally distributed and the covariance matrices for the variables are the same for both groups. It is likely that these assumptions are not optimal for our data, and this fact will reduce the effectiveness of any algorithm derived using linear methods. Discrimination techniques which do not make these assumptions and which are not restricted to linear functions of the variables (neural network methods for example) should be expected to provide superior classification speed and performance (Gorman and Sejnowski, 1988).

The discrimination procedure used in this paper also attempted to differentiate responses from all subjects using a single criterion. Another, possibly more successful approach would be to collect pilot data and develop a separate non-linear classification algorithm for each subject. We expect that a system sufficiently robust for routine use would need first to adapt to baseline data for each operator and then monitor several concurrent streams of physiological information using non-linear discriminant methods. It remains to be seen whether, in such a system combining multiple streams of electrophysiological information, measures of averaged or single-trial evoked responses can contribute enough predictive power to justify their use.

References

Beatty, J., Greenberg, A., Deibler, W.P., & O'Hanlon, J.F. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science* 183, 871-873.

Belyavin, A. & Wright, N.A. (1987). Changes in electrical activity of the brain with vigilance. *Electroenceph clin Neurophysiol* 66, 137-144.

Bente, D. (1964). *Die Insuffizienz des Vigilitatstonus*. Habilitationsschrift, Erlangen.

Fruhstorfer, H. & Bergstrom, R.M. (1969). Human vigilance and auditory evoked responses. *Electroenceph clin Neurophysiol* 27, 346-355.

Goeller, C.J. & Sinton, C.M. (1989). A microcomputer-based sleep stage analyzer. *Computer Methods and Programs in Biomedicine* 29, 31-36.

Gorman, R.P. & Sejnowski, T.J. (1988) Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks* 1, 75-89.

Hillyard, S.A., Hink, R.F., Schwent, V.L., & Picton, T.W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177-180

Kramer, A.F. (1991). Physiological metrics of mental workload: a review of recent progress. In: D. Damos (Ed.), *Multiple Task Performance*. New York: Taylor & Francis.

Loomis, A.L., Harvey, E., & Hobart, G.A. (1937). Cerebral states during sleep, as studied by human brain potentials. *J Exp Psychol* 21, 127-144.

Mackworth, J.F. (1970). *Vigilance and attention: A signal detection approach*. New York:

Penguin Books.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quart J Exp Psychol* 1, 6-21.

Makeig, S. & Inlow, M.H. (1991). Lapses in alertness: Coherence of fluctuations in performance and EEG spectrum. Unpublished technical report, Naval Health Research Center, San Diego.

Matousek, M. & Petersen, I. (1983). A method for assessing alertness fluctuations from EEG Spectra. *Electroenceph clin Neurophysiol* 55, 108-113.

Ogilvie, R.D., Simons, I.A., Kuderian, R.H., MacDonald, T., and Rustenburg, J. (1991). Behavioral, event-related potential, and EEG/FFT changes at sleep onset. *Psychophysiology* 28, 54-64.

Ornitz, E.M., Ritvo, E.R., Carr, E.M., La Franchi, S., & Walter, R.D. (1967). The effect of sleep onset on the auditory averaged evoked response. *Electroenceph clin Neurophysiol* 23, 35-341.

Penzel, T. & Petzold, J. (1989) A new method for the classification of subvigil stages, using the Fourier transform, and its application to sleep apnea. *J Comput Biol Med* 19, 7-34.

Rechtschaffen, A., and Kales, A., eds. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Public Health Service, Bethesda, Maryland.

Roth, B. (1961). The clinical and theoretical importance of EEG rhythms corresponding to states of lowered vigilance. *Electroenceph clin Neurophysiol* 13, 395-399.

Santamaria, J. & Chiappa, K.H. (1987). The EEG of drowsiness in normal adults. *J Clin Neurophysiolog* 4, 327-382.

Torsvall, L. & Akerstedt, T. (1988). Extreme sleepiness: quantification of EOG and spectral EEG parameters. *Int J Neurosci* 38, 435-441.

Townsend, R.E. and Johnson, L.C. (1979). Relation of frequency-analyzed EEG to monitoring behavior. *Electroenceph clin Neurophysiol* 47, 272-279.

Ulrich, G. & Frick, K. (1986). A new quantitative approach to the assessment of stages of vigilance as defined by spatiotemporal EEG patterning. *Percept Mot Skills* 62, 567-576.

Williams, H.L., Tepas, D.I., & Morlock, H.C. (1962). Evoked responses to clicks and electroencephalographic stages of sleep in man. *Science* 138, 685-686.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 01/92	3. REPORT TYPE AND DATE COVERED FINAL 06/89 - 09/90

4. TITLE AND SUBTITLE
Lapses in alertness: Brain-evoked responses to task-irrelevant auditory probes

5. FUNDING NUMBERS
Program Element:
Work Unit Number:
REIMB
NPRDC - 9225

6. AUTHOR(S)
Scott Makeig, F. Scot Elliott, Mark Inlow, David Kobus

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No.   90-39

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
N/A

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Thirteen subjects participated in an auditory simulation of a passive sonar target detection environment. Targets were 300 ms noise bursts presented at near threshold levels in a noise background at a mean rate of 10 per minute. Task-irrelevant probe tones were also presented at inter-stimulus intervals of 2-4 seconds. Each subjects participated in two 28 minute test sessions, pressing a button whenever they detected a noise target. Prominent minute-scale fluctuations in performance (computed as changes in local error rate using a 32-s moving window) occurred in many of the sessions. Evoked responses to the irrelevant probe tones in thirteen runs with highest number of performance lapses were sorted by current local error rate and smoothed using a moving-average. The amplitude of the grand mean N2 response to the irrelevant probe tones increased monotonically with error rate. Averaged evoked responses to relatively frequent, task-irrelevant probe tones appear to allow an accurate estimate of level of alertness if adequate numbers of trials are available.

14. SUBJECT TERMS
Vigilance alertness, sleep evoked response, attention auditory sonar

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500